```
## Selection models in R ##

install.packages("sampleSelection")
library(sampleSelection)

data("Mroz87")
Mroz87$kids <- (Mroz87$kids5 + Mroz87$kids618 > 0)
head(Mroz87)

# Female labor supply (lfp = labour force participation)

## Outcome equations without correcting for selection
# I() means "as-is" -- do calculation in parentheses then use as variable

## Comparison of linear regression and selection model

outcome1 <- lm(wage ~ exper, data = Mroz87)
summary(outcome1)

selection1 <- selection(selection = lfp ~ age + I(age^2) + faminc + kids + educ, outcome =
wage ~ exper,
data = Mroz87, method = "2step")
summary(selection1)

plot(Mroz87$wage ~ Mroz87$exper)
curve(outcome1$coeff[1] + outcome1$coeff[2]*x, col="black", lwd="2", add=TRUE)
curve(selection1$coeff[1] + selection1$coeff[2]*x, col="orange", lwd="2", add=TRUE)


## A more complete model comparison

outcome2 <- lm(wage ~ exper + I( exper^2 ) + educ + city, data = Mroz87)
summary(outcome1)

## Correcting for selection

selection.twostep2 <- selection(selection = lfp ~ age + I(age^2) + faminc + kids + educ,
outcome = wage ~ exper + I(exper^2) + educ + city,
data = Mroz87, method = "2step")
summary(selection.twostep2)

selection.mle <- selection(selection = lfp ~ age + I(age^2) + faminc + kids + educ, outcome =
wage ~ exper + I(exper^2) + educ + city,
data = Mroz87, method = "mle")
summary(selection.mle)


## Heckman model selection "by hand" ##

seleqn1 <- glm(lfp ~ age + I(age^2) + faminc + kids + educ, family=binomial(link="probit"),
data=Mroz87)
summary(seleqn1)

## Calculate inverse Mills ratio by hand ##

Mroz87$IMR <- dnorm(seleqn1$linear.predictors)/pnorm(seleqn1$linear.predictors)

## Outcome equation correcting for selection ##

outeqn1 <- lm(wage ~ exper + I(exper^2) + educ + city + IMR, data=Mroz87, subset=(lfp==1))
summary(outeqn1)

## compare to selection package -- coefficients right, se's wrong
summary(selection.twostep2)
```

```
## interpretation
## If our independent variables does not appear in the selection equation, we can interpret beta as in linear regression
## If it does appear in the selection equation, we must calculate:

beta.educ.sel <- selection.twostep2$coefficients[6]
beta.educ.out <- selection.twostep2$coefficients[10]
beta.IMR <- selection.twostep2$coefficients[12]
delta <- selection.twostep2$imrDelta

marginal.effect <- beta.educ.out - beta.educ.sel * beta.IMR * delta
mr2 <- marginal.effect * Mroz87$educ


plot(Mroz87$wage ~ Mroz87$educ)
lines(mr2 ~ Mroz87$educ, type="l", col="green", lwd="2")

## Selection with a binary outcome variable
## Data from Kimball (2006)

library(foreign)
conflict.data <- read.dta("MissingLink_JPRfinal.dta",  convert.factors=FALSE)
conflict.data <- na.omit(conflict.data)
head(conflict.data)

## A probit model for conflict

probit.model <- glm(conflict ~ relcap+contig+jtdem+jaut+pwrs+allform2,
family=binomial(link=probit), data=conflict.data)
summary(probit.model)

install.packages("Zelig")
library(Zelig)
install.packages("ZeligChoice")
library(ZeligChoice)


selection.formula <- list(mu1 = conflict ~ relcap+contig+jtdem+jaut+pwrs,
mu2 = allform2 ~ relcap+logdist+contig+jtdem+jaut+sharerival)

selection.binary <- zelig(selection.formula, model = "bprobit", data = conflict.data)
summary(selection.binary)

x.contig <- setx(selection.binary, contig=1)
sim.binary1 <- sim(selection.binary, x = x.contig)
summary(sim.binary1)
plot(sim.binary1)

x.noncontig <- setx(selection.binary, contig=0)
sim.binary2 <- sim(selection.binary, x = x.contig, x1=x.noncontig)
summary(sim.binary2)
plot(sim.binary2)
```